Supporting Information for Lowered pH Leads to Fusion Peptide Release and a Highly-dynamic Intermediate of Influenza hemagglutinin

Xingcheng Lin[1], Jeffrey K. Noel[1], Qinghua Wang[2], Jianpeng Ma[1,2,3], and José N. Onuchic*[1]

[1]Center for Theoretical Biological Physics, Rice University, Houston, TX
[2]Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX
[3]Department of Bioengineering, Rice University, Houston, TX
*Email: jonuchic@rice.edu, Phone: (713)348-4197

Contents

S1 Pre- and Post-fusion HA_2 crystal structure

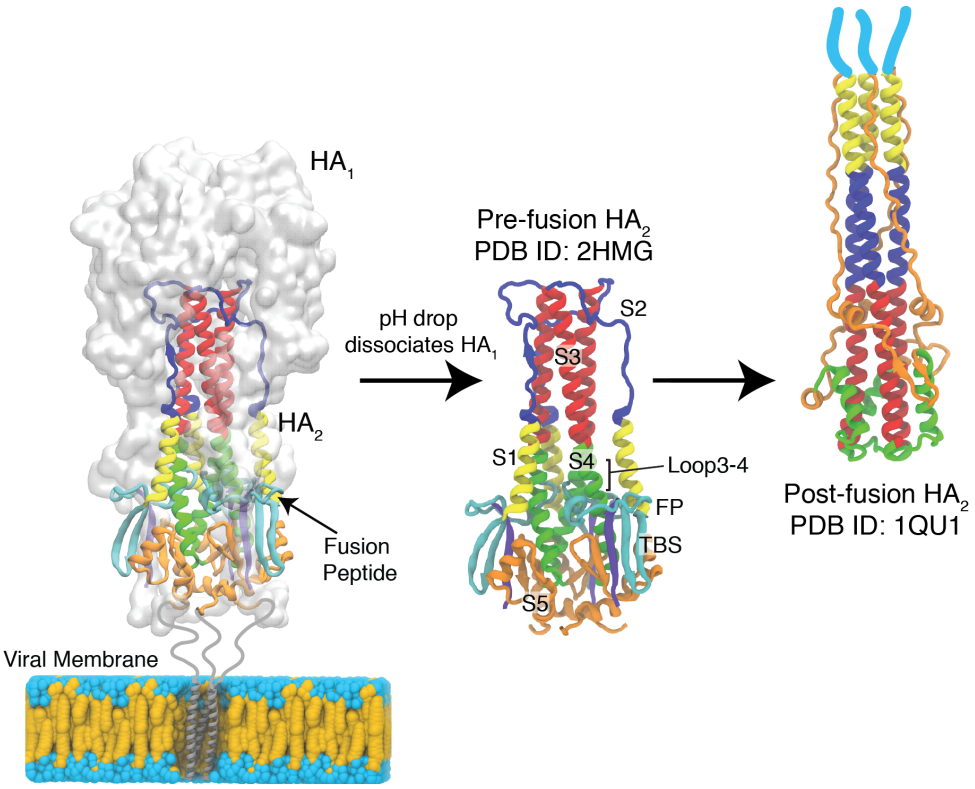

Figure S1: **(Left)** A complete HA including its receptor binding domain HA_1 and membrane fusion domain HA_2. The whole molecule is oriented with respect to a cartoon membrane. After endocytosis, the pH drop causes a dissociation of HA_1 from the HA_2 trimer and sterically allows the conformational transition of HA_2. **(Right)** Comparison between pre-fusion (PDB ID 2HMG) and post-fusion (PDB ID 1QU1) HA_2 crystal structures [1, 2, 3]. They are aligned by S3, the only domain structurally conserved in between these two structures. The fusion peptides (cyan, residue ID 1 to 20 in each monomer, abbreviated as FPs afterwards) are buried in a hydrophobic pocket in the pre-fusion structure. S1 (yellow, residue ID 38 to 54 in each monomer, also named A-helix) moves over 100 Å. S2 (blue, residue ID 55 to 75 in each monomer, also named B-loop) goes through a loop-to-helix transition. S3 (red, residue ID 76 to 105 in each monomer) is structurally conserved between these two structures. S4 (green, residue ID 106 to 129 in each monomer) reorients and packs antiparallel to S3. S5 (orange, residue ID 130 to 175 in each monomer) breaks from its pre-fusion globular compact structure into three loops flanking the grooves of the long coiled-coil. Two beta-strand region (cyan, residue ID 21 to 37 in each monomer, abbreviated as TBS afterwards) connects FPs and S1. Loop3-4 (residue ID 107 to 112) changes from helical into a loop connecting S3 and S4 in the post-fusion structure. In addition, HA_1 (three small magenta peptide fragments) are disulfide bonded to HA_2 (connecting residue ID 14 of HA_1 to residue ID 137 of HA_2). Two beta-strand region (TBS) and FPs are not available in the post-fusion crystal structure and thus are shown as three cyan lines. Figure are prepared with VMD [4].

S2 Detailed methods

Explicit solvent simulations were performed at either physiological temperature (T = 310K) or high temperature (T = 398K). The protein was solvated with the TIP3P water model, Na^+ and Cl^- ions were used to neutralize and mimic the endosomal ionic environment of HA molecule, with a physiological concentration of 0.15 M. The simulations were performed with a time step of 2 fs. The forcefield CHARMM22* [5] was used. CHARMM22* was created to correct the overstabilization of helices in CHARMM22/CMAP and to more accurately describe salt-bridge interactions [5].

S2.1 Physiological temperature simulations on Anton

Anton is a special machine designed for molecular dynamics simulation [6]. Before all production runs, the protein was equilibrated using Desmond in the NPT ensemble at 1 bar for 3 ns. Three different simulation setups were used to capture various stages of the HA_2 transition. Complete sets of simulations with physiological temperature (T = 310K) are listed at Table S1:

1. Ejection of FPs. Complete HA_2 including disulfide bonds to three HA_1 peptides (residue ID 10 to 18 of HA_1). The box was $76 \times 74 \times 119$ $Å^3$ and contained 68147 atoms. This simulation was done with both the forcefields of CHARMM22* and CHARMM36 (Table S1 Simulations 1-4)

2. FPs removed. Identical to 1. but with FPs (residue ID 1 to 20 of HA_2) deleted from the simulation box. The box was $82 \times 79 \times 123$ $Å^3$ with a total of 81268 atoms. (Table S1 Simulation 5)

3. FPs, S1, S2, and TBS removed. Starting from the last snapshot of the previous simulation, it was performed in a box of $103 \times 103 \times 117$ $Å^3$ with a total of 125540 atoms. (Table S1 Simulation 6)

S2.2 Results with CHARMM36

CHARMM22/CMAP has been criticized for being overly helical [7] and two different approaches were taken to correct it, CHARMM22* [5] and CHARMM36 [8]. Since the FPs are buried in the loop3-4 helical region, for robustness, we wanted to observe FP release in both these forcefields. In our analysis, CHARMM36 was more helical than CHARMM22*. Nonetheless, in two of the three simulations with CHARMM36, we observed the release of the first FP within 1μs. Note, that among all Set 1 simulations, each of the three FP fell off at least once (i.e. evidence of no asymmetry in the crystal structure, nor in our equilibration) (Fig. S2). We perform the same calculation for S3/S4 helical order in the simulations with CHARMM36 (Fig. S3) to compare to Fig. 2 in the main text. It shows that Loop3-4 is more helical with CHARMM36 than with CHARMM22*. Thus,

Simulation	Set[a]	Simulated system	Forcefield	Starting Topology	Length (μs)	Result
1	1	HA_2 +P1[b]	CHARMM22*	Crystal structure	2.0	FP falls off [c]
2	1	HA_2 +P1	CHARMM36	Crystal structure	2.0	FP falls off
3	1	HA_2 +P1	CHARMM36	Crystal structure	1.0	FP falls off
4	1	HA_2 +P1	CHARMM36	Crystal structure	1.0	No FP falls off
5	2	HA_2 (resID 21-175)	CHARMM22*	Snapshot from trajectory 1 where one FP falls off	2.0	Loop3-4 partially disorder
6	3	HA_2 (resID 76-175)	CHARMM22*	Snapshot from trajectory 5	6.0	Loop3-4 completely disorder and S3 bent towards S5

[a]The simulations are categorized as 3 sets
[b]P1: Residue ID 10 to 18 of HA_1 that is disulfide bonded to HA_2
[c]FP falls off means at least one FP falls off during simulation

Table S1: Summary of Physiological temperature (T = 310K) simulations in this paper

while the local helical disorder observed in CHARMM22* provides a possible mechanism for FP release, the CHARMM36 results indicate that the disorder is not a necessary condition for release.

S2.3 Estimation of Protonation Rate

To estimate the rate of protonation for each solvent accessible residues, we utilized a diffusion-limited Smoluchowski encounter model [10, 11, 12]. By assuming a perfect receptor, which means there is 100% chance for protons to protonate the relevant residue once it collides into it, the resulting rate of uptake will be a upper limit:

$$k = 4\pi DRc \tag{1}$$

where D is the diffusion coefficient of protons in water, R is the radius of one residue, c is the concentration of protons at a certain pH. Here, pH = 5 corresponds to c = $10^{-5}M = 6.02 \times 10^{-6}/nm^3$. We take the upper limit of the radius for one residue as R = 1 nm, an exaggeration since the real target of protonation (involving usually only one oxygen atom per residue) has a much smaller radius. The diffusion coefficient D, considering the Grotthuss mechnism[13], is around $7 \times 10^{-5} cm^2/s$ [14]. Taking into account all of these parameters, the resulting estimation for the upper limit of protonation rate is $k \approx 5 \times 10^5/s$, i.e., the average time for one residue to be protonated is about $2\mu s$. Based on that, we estimate in the ideal case, the proton transfer at pH 5 will take at least 1-5 μs.

S2.4 High temperature simulations

The dissociation of FPs was too slow at 310 K to generate statistics when the FP pocket was stabilized by salt bridges. Thus, higher temperature was used to speed up the barrier crossing [15]. Often these type of simulations are performed at around 500K, but here a lower temperature of 398K was sufficient to realize the kinetic events within our computational capabilities. Even if no FP dissociated, a simulation was stopped after a maximum of 200 ns.

To test various protonation states, four sets of high temperature simulations were implemented. The full HA_2 is used, as described in simulation Set 1 above. Assignment of protonation states is described in the next section.

1. 10 simulations at pH 4.5 plus the three ASP_{112} of HA_2 were protonated (to break salt bridge with the FPs).

2. 6 simulations similar to the first ones except that only one ASP_{112} of HA_2 was protonated (to break salt bridge with its corresponding FP).

3. 7 simulations at pH 4.5 without protonating the three ASP_{112}. This means all ASP_{112} remained charged and able to form their respective salt bridges.

4. 6 simulations at pH 7.0, but with the three ASP_{112} protonated. This provided a control as to the effect of the overall charge state of HA_2 on the FPs.

In detail, for simulation set 1 (Table S1), a preliminary simulation with residues titrated at pH 4.5 according to [16] was performed for 1μs started from crystal structure. Then the PROPKA algorithm was applied to each snaphot to calculate the pKa of all titratable residues during this simulation. Thereafter, the average charge states of titratable residues were fixed. For example, ASP_{109} was shown to be neutral in more than 500 ns of this 1μs simulation, so it was neutralized during the production run. In set 2, the three FP monomers were removed, modeling their dissociation. To best model the change, we recalculated the average charge states of the residues after FP ejection in simulation set 1. Hereafter, pH 5.0 (the likely lowest pH reached in vivo) was used to better describe the endosomal environment. In set 3 we break the S1-S4 interface by deleting S1, S2, TBS and its attached HA_1 peptide. Thus, we recalculate the protonation states from a group of Set 2 structures with these N-terminal pieces deleted.

[1]Label notation: Histidine at chain 1 with residue ID 17
[2]Label notation: + Protonated - Not protonated NA Not applicable

Table S2: The protonation details of titratable residues in the physiological temperature simulation. All ARG and LYS are positively charged during the simulations.

Residues	Set 1 (pH=4.5)	Set 2 (pH=5.0)	Set 3 (pH=5.0)
$HIS17_1$ [1]	+ [2]	+	NA
$HIS18_1$	+	+	NA
$HIS26_2$	+	+	NA
$HIS64_2$	+	+	NA
$HIS106_2$	+	+	+
$HIS142_2$	+	+	+
$HIS159_2$	+	+	+
$ASP19_2$	-	NA	NA
$ASP37_2$	-	-	NA
$ASP46_2$	-	-	NA
$ASP79_2$	-	-	-
$ASP86_2$	-	-	-
$ASP90_2$	-	-	-
$ASP109_2$	+	+	-
$ASP112_2$	+	+	-
$ASP145_2$	-	-	-
$ASP158_2$	-	-	-
$ASP160_2$	-	-	-
$ASP164_2$	-	-	-
$GLU11_2$	+	NA	NA
$GLU15_2$	+	NA	NA
$GLU30_2$	+	-	NA
$GLU57_2$	+	-	NA
$GLU61_2$	+	-	NA
$GLU67_2$	+	-	NA
$GLU69_2$	+	-	NA
$GLU72_2$	+	-	NA
$GLU74_2$	+	-	NA
$GLU81_2$	+	+	-
$GLU85_2$	-	-	-
$GLU97_2$	+	+	-
$GLU103_2$	-	-	-
$GLU114_2$	+	+	-
$GLU120_2$	+	+	-
$GLU128_2$	+	+	+
$GLU131_2$	+	-	-
$GLU132_2$	-	-	-
$GLU150_2$	+	-	-
$GLU165_2$	+	-	-

S2.5 Definition of θ_{S3}

To quantify the observed bending of S3 towards S5 (which we term as "symmetry breaking"), we compare the positions of S3 and S5 relative to the crystal structure alignment. Parallel central axes of S3 and S5, perpendicular to the viral surface were defined based on the crystal coordinates. The crystal structure coordinates of S3 or S5 were aligned by least RMSD to each simulation snapshot S3 and S5 respectively. Then the bending angle θ_{S3} could be calculated by measuring the internal angle between the central axes of these two aligned

domains (Main text Fig. 3).

S2.6 Calculation of the effective diffusion coefficient of θ_{S3}

The mean square deviation of the bending angle θ_{S3} behaves linearly as a function of time at long times (main text Fig. 3D), indicating a diffusive behavior. The effective diffusion coefficient along this reaction coordinate can be calculated as [17]:

$$D = \lim_{\tau \to \tau_D} \frac{1}{2} \frac{\partial}{\partial \tau} \langle \delta\theta_{S3}(\tau)^2 \rangle = 7.33 \times 10^2 \text{degree}^2/\mu\text{s} \tag{2}$$

where τ_D is the time interval where $\langle \delta\theta_{S3}(\tau)^2 \rangle$ becomes linear.

S3 FPs release after S5 breaks apart when ASP_{112} is negatively charged

In the high temperature simulations where ASP_{112} was negatively charged, the salt-bridge interaction between ASP_{112} and the N-terminus of FP was intact. In this case, FPs were more stable buried in their pre-fusion location than the stability of the domain S5. We quantify this with a measure $Q_{S5-\text{interface}}$, the proportion of S5 inter-monomer native contacts formed. Figure S4 shows $Q_{S5-\text{interface}}$ and the distance between FP and its buried hydrophobic pocket, for a typical high temperature trajectory. It is clear that S5 breaks ($Q_{S5-\text{interface}}$ close to 0) before the release of any FP (distance larger than 2.5 nm).

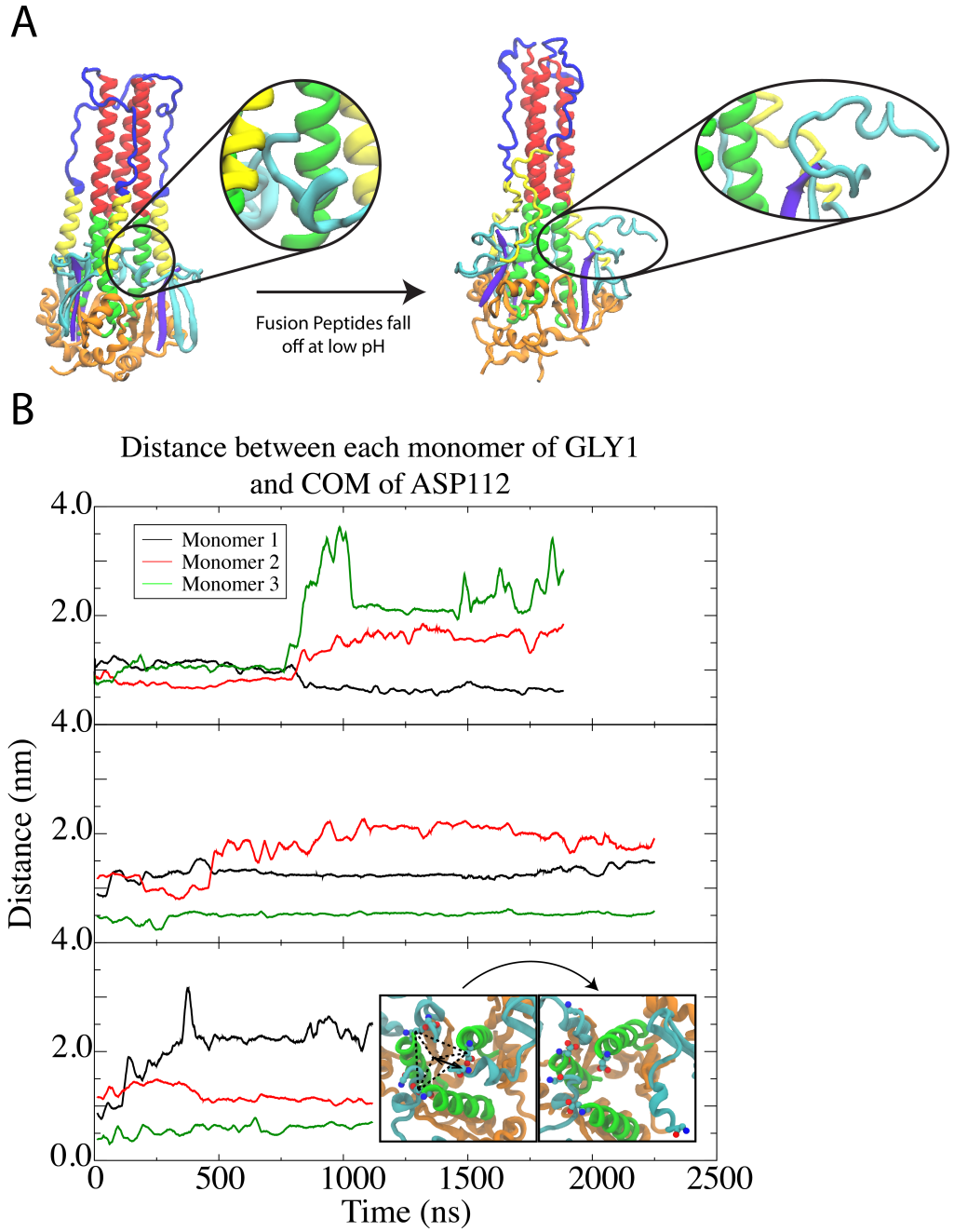

Figure S2: The simulation under physiological temperature (T = 310K) started from HA_2 crystal structure. A. The sideview crystal structure before (left) and after (right) FPs falling off (as compared to the top-view shown in the main text). B. The distance between the N-terminus of each of FPs monomers and the center of hydrophobic pocket where FPs are buried in the pre-fusion structure, represented as the center of mass of residue ASP_{112}. The curves are averaged over a 0.5 ns window for clarity. The (magnified) inset at the bottom defines the geometric center of the hydrophobic pocket as center of mass of the three ASP_{112} in the crystal. The top plot represents the same simulation with CHARMM22* forcefield shown in the main text. The other two plots were from the simulations with CHARMM36 forcefield. In all three plots, FPs fell off eventually. Note three different FP monomer (represented by different colors) fell off firstly in these simulations.

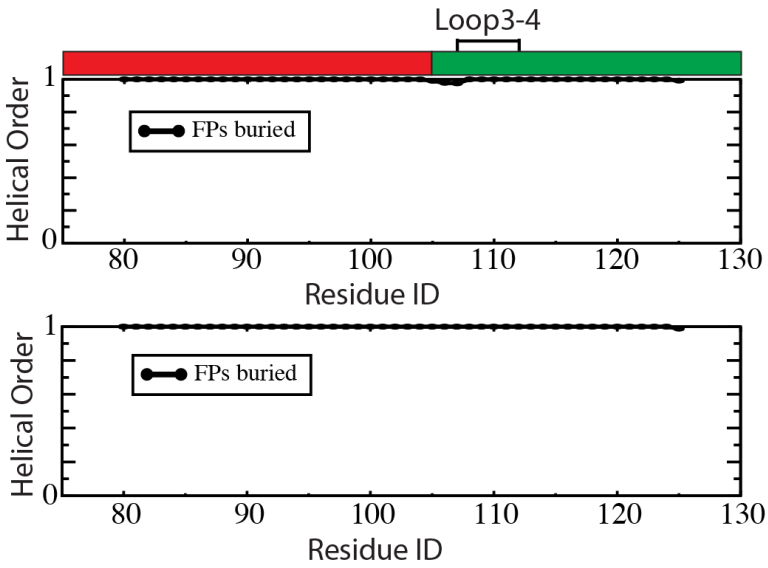

Figure S3: Helical order of domains S3/S4 in the first set of the simulations with CHARMM36 forcefield. The helical order of a residue determined by averaging over the three monomers the probability that the STRIDE algorithm [9] determines an alpha helical secondary structure. Hence here 0 means completely disordered and 1 means perfectly ordered during the calculation period. Overall the residues under CHARMM36 forcefield are more helical than those under CHARMM22* forcefield.

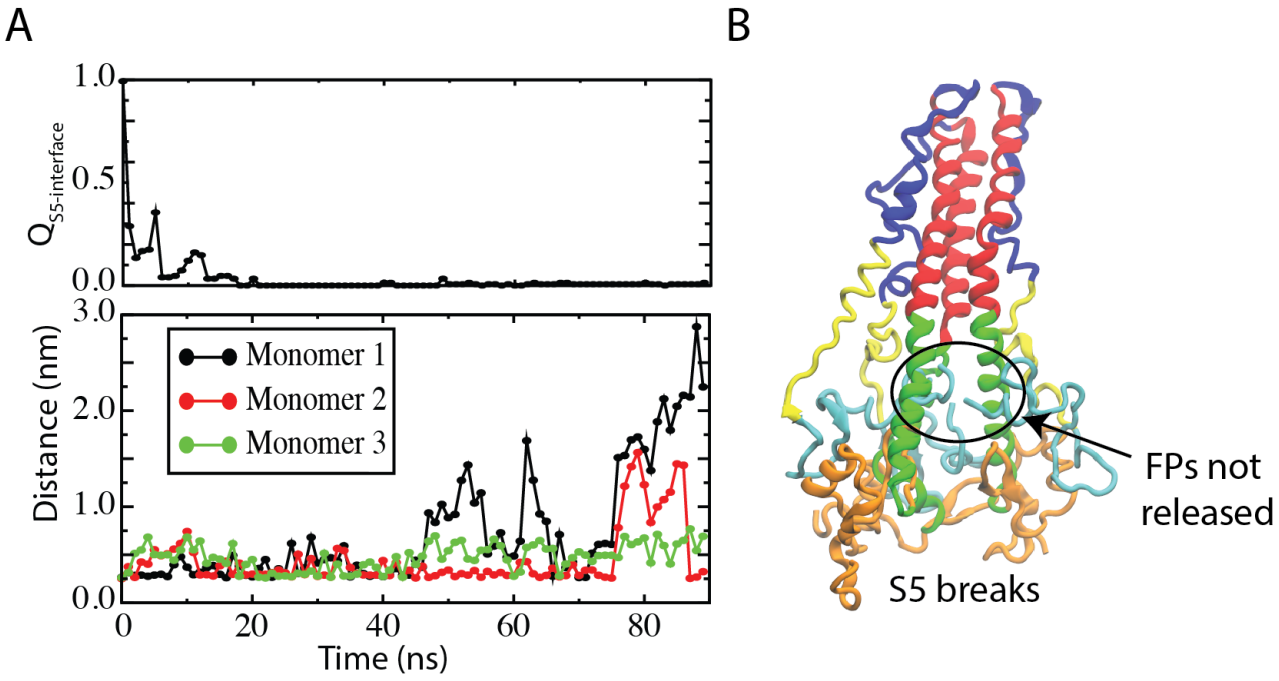

Figure S4: Quantitative comparison between the release of FPs and the breaking of S5. A. Top: The change of QS5-interface, the proportion of inter-monomer native contacts formed in the pre-fusion S5 domain. Bottom: The change of distance between each of the FPs and the center of the hydrophobic pocket where FPs are buried in the pre-fusion structure. S5 breaks (QS5-interface close to 0) before the release of any FP (distance larger than 2.5 nm). B. One snapshot taken from the high temperature simulation showing the S5 breaks while all FPs are still close to its hydrophobic pocket.

References

[1] Wilson, I. A, Skehel, J. J, & Wiley, D. C. (1981) Structure of the hemagglutinin membrane glycoprotein of influenza-virus at 3-a resolution. *Nature* **289**, 366–373.

[2] Bullough, P. A, Hughson, F. M, Skehel, J. J, & Wiley, D. C. (1994) Structure of influenza hemagglutinin at the ph of membrane-fusion. *Nature* **371**, 37–43.

[3] Lin, X, Eddy, N. R, Noel, J. K, Whitford, P. C, Wang, Q, Ma, J, & Onuchic, J. N. (2014) Order and disorder control the functional rearrangement of influenza hemagglutinin. *Proc. Natl. Acad. Sci. U. S. A* **111**, 12049–12054.

[4] Humphrey, W, Dalke, A, & Schulten, K. (1996) Vmd: Visual molecular dynamics. *Journal of Molecular Graphics and Modelling* **14**, 33–38.

[5] Piana, S, Lindorff-Larsen, K, & Shaw, D. E. (2011) How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* **100**, L47–L49.

[6] Shaw, D. E, Deneroff, M. M, Dror, R. O, Kuskin, J. S, Larson, R. H, Salmon, J. K, Young, C, Batson, B, Bowers, K. J, Chao, J. C, Eastwood, M. P, Gagliardo, J, Grossman, J. P, Ho, C. R, Ierardi, D. J, Kolossvary, I, Klepeis, J. L, Layman, T, McLeavey, C, Moraes, M. A, Mueller, R, Priest, E. C, Shan, Y, Spengler, J, Theobald, M, Towles, B, & Wang, S. C. (2008) Anton, a special-purpose machine for molecular dynamics simulation. *Communications of the Acm* **51**, 91–97.

[7] Best, R. B, Buchete, N. V, & Hummer, G. (2008) Are current molecular dynamics force fields too helical? *Biophys. J.* **95**, L7–L9.

[8] Best, R. B, Zhu, X, Shim, J, Lopes, P. E. M, Mittal, J, Feig, M, & MacKerell, Alexander D., J. (2012) Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *Journal of Chemical Theory and Computation* **8**, 3257–3273.

[9] Frishman, D & Argos, P. (1995) Knowledge-based protein secondary structure assignment. *Proteins-Structure Function and Genetics* **23**, 566–579.

[10] Smoluchowski, M. V. (1917) Versuch einer mathematischen theorie der koagulationskinetik kolloider lsungen. *Z. Phys. Chem* **92**, 129–168.

[11] Dorsaz, N, De Michele, C, Piazza, F, De Los Rios, P, & Foffi, G. (2010) Diffusion-limited reactions in crowded environments. *Phys. Rev. Lett.* **105**.

[12] Phillips, R, Kondev, J, J, T, Garcia, H, & Orme, N. (2013) Chapter 13 a statistical view of biological dynamics. *Physical Biology of the Cell*.

[13] de Grotthuss, C. J. T. (1806) Sur la dcomposition de l'eau et des corps qu'elle tient en dissolution a l'aide de l'lectricit galvanique. *Ann. Chim.* **58**, 54–74.

[14] Agmon, N. (1995) The grotthuss mechanism. *Chemical Physics Letters* **244**, 456–462.

[15] Day, R, Bennion, B. J, Ham, S, & Daggett, V. (2002) Increasing temperature accelerates protein unfolding without changing the pathway of unfolding. *J. Mol. Biol.* **322**, 189–203.

[16] Zhou, Y, Wu, C, Zhao, L, & Huang, N. (2014) Exploring the early stages of the ph-induced conformational change of influenza hemagglutinin. *Proteins-Structure Function and bioi* **82**, 2412–2428.

[17] Yeh, I. C & Hummer, G. (2004) Diffusion and electrophoretic mobility of single-stranded rna from molecular dynamics simulations. *Biophys. J.* **86**, 681–689.